DOCUMENT RESUME

ED 095 210 TM 003 883

| | |
|---|---|
| AUTHOR | Gilman, David A. |
| TITLE | Alternatives to Tests, Marks, and Class Rank. |
| INSTITUTION | Indiana State Univ., Terre Haute. Curriculum Research and Development Center. |
| PUB DATE | May 74 |
| NOTE | 56p. |
| AVAILABLE FROM | Curriculum Research and Development Center, Jamison Hall, School of Education, Indiana State University, Terre Haute, Indiana 47809 ($1.00) |
| EDRS PRICE | MF-$0.75 HC-$3.15 PLUS POSTAGE |
| DESCRIPTORS | Criterion Referenced Tests; Decision Making; *Evaluation Techniques; Grade Point Average; *Grading; Norm Referenced Tests; Pass Fail Grading; Rating Scales; Self Evaluation; *Student Evaluation |
| IDENTIFIERS | Class Rank; *Contract Grading; Performance Based Evaluation; Self Scoring Tests |

ABSTRACT

The author reviews various current approaches to evaluating pupil achievement, evaluating instruction, and using grades, grade point averages, and class rank. Such issues as the philosophical and psychological bases for selection and use of evaluation instruments, the use of performance objectives, ways to reduce subjectivity in grading, the use of self-scoring tests, criterion referenced measurement, contract grading, and self-evaluation are discussed with the practitioner in mind. (Author/RC)

ED 095210

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY.

BEST COPY AVAILABLE

# ALTERNATIVES TO TESTS, MARKS, AND CLASS RANK

CURRICULUM RESEARCH
AND DEVELOPMENT CENTER

SCHOOL OF EDUCATION, INDIANA STATE UNIVERSITY ● TERRE HAUTE

## THE CURRICULUM RESEARCH AND DEVELOPMENT CENTER

### School of Education, Indiana State University

The Curriculum Research and Development Center of Indiana State University provides school systems the opportunity to secure aid, encouragement, and cooperation in curriculum development projects. It coordinates the participation of University personnel engaged in curriculum work, provides information concerning curriculum development, and initiates and sponsors curriculum research projects. It is the contact point where public schools initiate inquiries regarding curriculum and acts as a vehicle for communication between elementary and secondary schools and the University. Although the CRDC operates as an agency of the School of Education, it represents all departments of the University engaged in curriculum development projects.

Services of the Center are of these major types:

Consultant services—The Center makes available specialists in various curriculum areas to aid in curriculum development. This is on a contract basis and can be adjusted to the needs of the particular task.

Leadership—In the Center, University personnel and others have an agency through which curriculum change can be advocated and other opportunities to exercise leadership provided.

Workshop and Conferences—One function of the Curriculum Research and Development Center is the planning and directing of on-campus conferences and workshops as well as those carried on in a school system engaged in curriculum development.

Printed Materials—The Center produces courses of study, bibliographies, and similar materials.

Research and Experimentation—The Center provides an environment which encourages research through supplying materials and services needed for experimentation, providing an atmosphere for communication and sharing of ideas, and aiding in acquiring funds for worthy projects.

David T. Turney
Dean, School of Education

Charles D. Hopkins
Director

The material for this bulletin was written by:

David A. Gilman
Professor of Education
Indiana State University
Terre Haute, Indiana

May 1974

## FOREWORD

During the past decade, educators have focused much attention on measurement and evaluation. This attention has been, at least in part, due to the general public's demand for what has been termed accountability. National and state assessment programs of educational progress have been or are being conducted focusing public attention on the outputs of education. One result of these assessment projects is the stimulation of educators to look for alternative approaches to testing and grading.

In this monograph, Professor Gilman reviews various current approaches to evaluating pupil achievement, evaluating instruction, and using grades, grade point averages, and class rank. Such issues as the philosophical and psychological bases for selection and use of evaluation instruments, the use of performance objectives, ways to reduce subjectivity in grading, the use of self-scoring tests, and many others are discussed in a manner that the practitioner at all levels of education--elementary, secondary, and college--will find equally applicable.

The controversy over the various approaches to educational measurement will in all probability continue; however, with the publication of this monograph the educational practitioner will be better able to understand some of the issues involved.

John C. Hill
Assistant Dean
Research and Services
Indiana State University

TABLE OF CONTENTS

INTRODUCTION 1

Accountability 1 Existentialism 1
Accountability vs. Existentialism 2 A Compromise 2

1. ALTERNATIVES TO TESTING: CRITERION REFERENCED MEASUREMENT 5

A Double Criterion 8 Steps in constructing CRM Tests 8
Differences in the Methods of NRM and CRM 11 Advantages 14
Limitations 14

2. ALTERNATIVES TO TESTING: SELF-SCORING TESTS 17

Varieties of Self-scoring Tests 18

3. ALTERNATIVES TO TESTING: THE CONTRACT METHOD 21

4. ALTERNATIVES TO GRADING 29

Arguments For and Against Grading 29 Emphasizing Strengths and Eliminating Weaknesses 32 Alternative Methods 33

5. ALTERNATIVE METHODS OF EVALUATION 37

Checklists and Rating Scales 37 Self-Evaluation and Self-Grading 41 Self-Evaluation 41 Self-grading 43 Pass-Fail 44 Equal Grading 45 Performance Based Evaluation 46

6. ALTERNATIVES TO GRADE POINT AVERAGE AND CLASS RANK 49

Reasons for Computing G.P.A. and Class Rank 49 Determination of Recipients of Scholarships and Financial Aid 50 Criticisms of Ranking and G.P.A. 51 Alternatives to G.P.A. and Class Rank 51 Summary 52

## LIST OF FIGURES


| | | |
|---|---|---|
| 1-1 | Scores on a Criterion Referenced Test | 5 |
| 1-2 | A Distribution of Test Scores with a Mean of 90 and a Standard Deviation of 10 | 6 |
| 1-3 | Steps in CRM (Model) | 9 |
| 1-4 | Steps in CRM (Listed) | 10 |
| 1-5 | Decision Making Process in NRM and CRM | 11 |
| 3-1 | A Whole Class Evaluation Contract | 22 |
| 3-2 | A Contract for an Individual Student | 24 |
| 3-3 | An Evaluation Schedule | 27 |
| 4-1 | A Computer Printout Narrative Report | 35 |
| 5-1 | An Evaluation Checklist | 38 |
| 5-2 | A Student Rating Scale | 39 |
| 5-3 | A Student-Teacher Rating Scale | 40 |
| 5-4 | A Self-Evaluation Form | 42 |
| 5-5 | A Performance Based Evaluation Report | 47 |

BEST COPY AVAILABLE

# ALTERNATIVES TO TESTS, MARKS, AND CLASS RANK

## INTRODUCTION

During the 1970's, few educational issues have stimulated such a diversity in opinion as the evaluation controversy.

There are growing numbers of individuals who are formulating measurement philosophies that are poles apart. Two camps with respect to testing philosophy are forming. The attitudes of these two groups concerning the importance of educational measurement could hardly be further apart.

### Accountability

First, a trend has developed that has been partly stimulated by advocates of educational accountability and has also been promoted by individuals concerned with evaluating the effectiveness of instruction. Associated with this trend are numerous educators who believe that if instruction is to be effective, the results of this effectiveness must be demonstrated. In many cases, the demonstration of this effectiveness has taken the form of comparing students' scores on standardized tests with national and local norms. Advocates of this approach to education cite evidence to show that education is the only industry that traditionally has never attempted to evaluate its products.

Advocates of accountability feel that testing and evaluation require much more attention in schools today than they are now receiving. They also believe that even if there were not a need to eliminate teachers who have a record of unsuccessful teaching attempts, teachers themselves should be continually trying to improve their instructional methods and in order to accomplish this, all teachers must engage in a rigorous effort to measure the results of instruction to determine if they have actually taught their students anything.

### Existentialism

Educators with an existentialist philosophy for many years have been concerned with the "humaneness" of education. The existentialists have long argued that the student is much more important in schools than learning is. They cite examples of the negative effect on students of competition arising from testing programs in the school. Frequently, anxiety, destructive competition, and emphasis on grades instead of learning are considered to be undesirable side effects of educational measurement. Many existentialists do not consider testing, grading, and ranking to be advantageous or even necessary evils. Some have gone so far as to say that measurement and evaluation of students are bad and ought to be discontinued. These points of view have often been associated with progressive education and a recurrence of these ideas has developed in recent years.

## Accountability vs. Existentialism

Most advocates of accountability look upon such "humane" approaches as elimination of testing, grading and ranking as a lot of sentimental tommyrot. They compile long lists of benefits they allege students receive from a well-run testing program. They believe that anything that exists and has value can be defined, measured, and evaluated in some form.

The existentialists believe that accountability is a symptom that was characteristic of education during the time of the "cult of efficiency." They believe that there are many things that are important in education, such as motivation, creativity, and attitudes, which cannot be measured by pencil and paper tests. They frequently see the emphasis on accountability as threatening to the education profession and to students because of the attention given quantities that are tangible and measurable.

## A Compromise

It is often possible to measure the effectiveness of instruction and yet avoid the negative effects that come about as a result of the side effects of testing. This booklet will attempt to describe and demonstrate some of the alternatives to testing that have been attempted and have been proven to be successful. Succinctly stated, this booklet will describe methods that can be used by teachers to determine the effectiveness of their instruction. These methods require neither competition between students nor emphasis on grades at the expense of learning. Furthermore, it is anticipated that teachers who evaluate students by any of these techniques will observe considerably fewer anxious students than they would with traditional grading practices.

This booklet is intended to be a guide for the educator rather than a research document. No attempt will be made to recount the historical background of these procedures. No attempt will be made to document literature sources to substantiate statements made concerning alternatives to testing. Persons interested in learning more concerning alternatives to testing or in investigating the merits of any particular alternative are urged to consult the *Education Index* or the ERIC files.

The author does not wish to become involved in the controversy between the advocates of accountability and the existentialists. Many good things do come from a well-run testing program. However, many of these benefits are also available with none of the alleged side effects caused by student-to-student competition.

Perhaps a word of caution should be offered to those who propose to use any of the alternatives in this booklet. It has been the author's experience that whenever an educator sees no need for testing or grading, he is frequently the very person who sets up an evaluation system that is bizarre, unfair to the students, and often less humane than the traditional methods of evaluation. Teachers who attempt any of the

systems described in this booklet are advised to use sound judgment and common sense in administering it. Educators who initiate one of the alternatives will be well advised to be sensitive to student feedback and to be flexible enough to change components of the system that are not producing the desired results.

The advantages as well as the disadvantages of each of the methods are described and suggestions are provided that will hopefully be helpful to those attempting an alternative approach to testing.

## Chapter 1

## ALTERNATIVES TO TESTING: CRITERION REFERENCED MEASUREMENT

One approach to the assessment of student achievement and the evaluation of instructional effectiveness that has been receiving a great amount of attention from educators who try to place a greater emphasis on individualized instruction is criterion referenced measurement (CRM).

In order to explain the criterion referenced method, it is necessary to contrast it with the traditional type of testing, which is known as norm referenced measurement (NRM).

A criterion is defined for purposes of CRM to be a standard of performance which serves as a minimum level to be used in a decision-making process. If a secretary is to be hired if, and only if, she can type 60 words per minute, then the ability to type 60 words per minute serves as the criterion for her employment as a secretary.

In Figure 1-1, the minimum standard of acceptable performance (the criterion) is that the student answer 90% of the items correctly. Student P answered 95% of the items correctly. Since his score is above criterion, he passes the test. Student F answered only 75% of the items correctly. His score is below criterion and he did not pass the test.

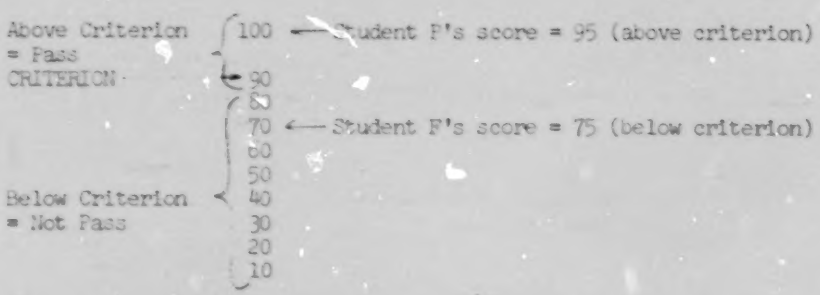


**Figure 1-1**
**Scores on a Criterion Referenced Test**

A norm may be thought of as an average. The mean, median, and mode are all examples of norms. Some of the types of scores derived from norm referenced information are percentiles, grade equivalent scores, age equivalent scores, I.Q. scores, standard scores, and stanine scores. To obtain these types of scores for a student, it is necessary to obtain

the mean or another type of average of the group the student belongs to. Frequently the relative distance a student scores from the mean is measured in units of standard deviations. A standard score of -1.0 means the student's score is one standard deviation below the mean while a standard score of +2.0 indicates the student's score is two standard deviations above the mean. (See Figure 1-2.)

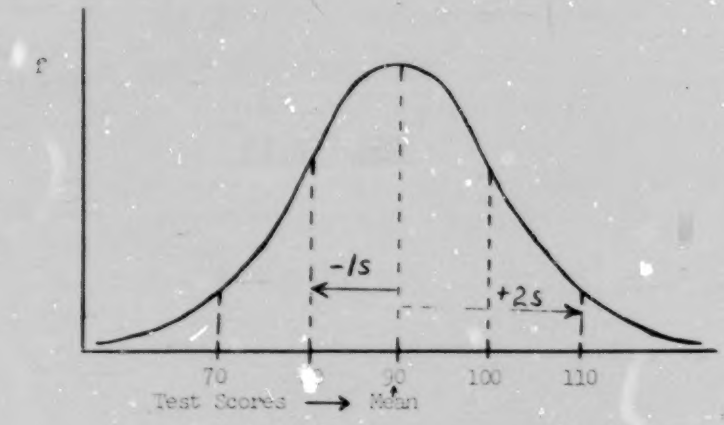


**Figure 1-2**
**A Distribution of Test Scores with a Mean of 90 and a Standard Deviation of 10**

Norm referenced measures are used to find out how each individual learner performs in relationship to the performance of other individuals on the same test. The only meaning the score has derives from its comparison with other scores and consequently with its comparison to the norm or average of the group. Each learner's performance is compared with the average students in his group (these measures are norm referenced measures). Most classroom tests and most standardized intelligence or achievements tests are norm referenced measures.

Criterion referenced measurement is one example of what can be called an absolute form of testing. Absolute interpretation of test scores involves making a judgment about the score of a student in terms of how his performance on the test relates to a certain standard or standards for test tasks.

Absolute interpretation of test performance is, of course, different from the traditional type of interpretation utilized in relative interpretations whereby the judgments about students' scores are based on the scores of other students in the group of which the

students are members.

Traditionally, testing experts, test theorists, and psychometric practitioners have given little attention to absolute interpretation of test performance. However, recently a great amount of attention has been devoted to this variety of testing by educational practitioners in a variety of areas.

An absolute interpretation of test scores is advocated and emphasized in such diverse fields as individualized instruction, programmed instruction, computer assisted instruction, non-graded schools, governmental education, industrial education, instructional technology, the systems approach to education, the British open school, military training, and physical education.

An important variety of measurement which requires an absolute interpretation of test scores, criterion referenced measurement was developed to be used as a technique for the assessment of a specific criterion behavior described in a behavioral instructional objective. This type of measurement focuses on the learning of an individual student at a particular point in time rather than on his standing among peers.

Criterion referenced measures focus attention on whether students are able to do certain tasks acceptably. It is because the learner is being compared to some established criterion, rather than to other individuals, that these measures are described as criterion referenced. The meaningfulness of any learner's score is *not* dependent on any comparison to the scores of other learners.

Since the student's anticipated performance as a result of learning has already been specified in a behavioral instructional objective, it is usually a very simple procedure to find out if the student can perform the behaviors specified in the objective. In many cases, learning or its absence may be demonstrated by having a student attempt the performance of a single act. However, it is more common for a test to be composed of a few items, rather than just one single item. In CRM, each test item is keyed to a set of behavioral objectives. CRM is designed to yield information directly relevant to the level or quality of behaviors that the examinee is capable of performing.

Although the results of CRM are interpretable in terms of the specified performance standards stated as the criteria of behavioral objectives, the level of these standards must be designated by the test constructor with full realization of the ability of his students and the importance of the behavior they are required to demonstrate.

An airline pilot will be expected to perform flawlessly on tests designed to measure his ability. A bright fourth grader may be expected to master *all* of the 100 multiplication facts. However, a social studies teacher may expect her slow learning students to obtain a score of at least 60% on the semester test. Consequently, she sets the criterion

level at 60%. A general education course taught at the college level may be taught in such a way that the instructor will consider his students as having mastered the material if they score higher than the criterion of 75%.

A frequently recommended criterion level is 90%. When a teacher sets up behavioral objectives for her class, the instruction and the CRM exercise are designed and constructed in a way that defines explicit rules linking patterns of test performance to behavioral objectives. If 90% is the criterion score, then any student who scores above 90% will be considered by the teacher to have learned the material. Students who score lower than 90% are considered to be below the desired level of mastery.

## A Double Criterion

Many instructors also use CRM to enable them to ascertain whether they are doing an effective job in teaching their classes by specifying a double criterion. The double criterion specifies the level of performance expected by each student in the class, and also specifies the number of students that should meet this standard in order for the instructor to consider that the instruction was successful.

An example of the double criterion can be found in the 90-90 criterion frequently used by authors of programmed instruction material. The 90-90 criterion means that the author may consider his work to be effective if 90% of the students are able to obtain a score at least equal to the criterion of 90% on the final examination. Any student who scores above 90% will be considered as having satisfactorily mastered the material. If 90% or more of the students score above this minimum level, the instructional materials are considered to be effective.

The choice of the level for the criterion or of the levels for the double criterion is determined by the instructor and is based on the competency of his students, the importance of the task, and the level of the instructor's aspiration. Some military training exercises specify a 95-95 criterion. However, for many teachers who wish to attempt CRM, a reasonable and challenging goal for any teacher is the 90-90 criterion.

A criterion referenced summative test is one that is deliberately constructed to give scores that tell what behaviors individuals with those scores have mastered. The standard (or criterion) against which a student's performance is compared represents the minimum acceptable performance for the desired behavior. Scores on the test for the entire class may be used to evaluate the effectiveness of instruction.

## Steps in Constructing CRM Tests

The step by step procedure for utilizing CRM is a logical, rational procedure. Some educators feel that to follow the sequence of steps required for the construction of a CRM instruments virtually guarantees an instructor that he will be effective.

The steps are as follows. Before instruction begins and before the test is constructed, the desired behaviors are carefully specified as instructional objectives. Then situations are created in which these desired behaviors may be demonstrated. Representative samples of the situations are then selected to test the tasks the learner is to perform. These sample situations constitute the CRM instrument. Next, instruction is planned so as to accomplish the instructional objectives. After the instruction has been completed, the CRM instrument is administered in an attempt to find (1) which students mastered the material as demonstrated by their above criterion scores and (2) if instruction was adequately effective as indicated by the percentage of students who attained the criterion score. The first of these functions of CRM is known as *summative evaluation*. The second of these functions, which represents an attempt to find whether instructional improvement is necessary, is known as *formative* evaluation.

Although the above sequence represents a rough sequential pattern of what occurs in CRM, perhaps Figure 1-3 and Figure 1-4 represent a more practical analysis of the sequence of CRM.

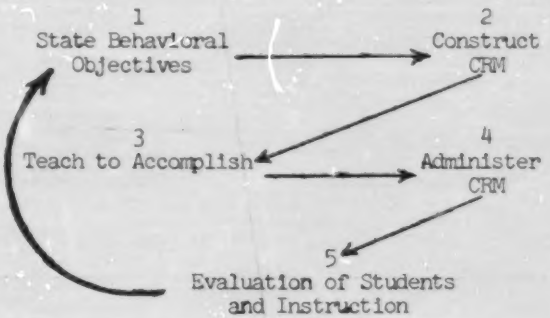


**Figure 1-3**
**Steps in CRM**

Figure 1-3 represents a model of what actually occurs in CRM. First, the instructional objectives are stated, preferably in the form of behavioral objectives.

Next, the measurement instrument is constructed in such a manner as to determine if the student can demonstrate the accomplishment of the behaviors described in the instructional objectives. The number of items the test will contain is up to the teacher. One guideline that may serve beginning CRM constructors is that it is well to have at least two items for each behavioral objective.

It is interesting to note that in the sequence of CRM, test construction is the second step, rather than next to last, as in the

construction of teacher-made NRM instruments.

The instruction is then performed so as to accomplish the objectives. Some critics of CRM have found fault with this step in the procedure, by alleging that at this point, the instructor is "teaching the test." It is a matter of individual perception as to whether this is happening or whether the objectives are truly being taught rather than the test. It is also a matter of debate as to whether there is something inherently wrong about teaching to the items on a test. Some advocates of CRM advise teachers to make students aware of their objectives so as to let the students know what will be expected of them in the evaluation procedure.

After the instruction is completed, the CRM instrument is administered and scored. There are only two possible scores. Students who score above the criterion pass, and those who score lower than criterion do not pass.

The scores of all students are then evaluated to determine if the instruction was effective. If the desired percentage of students attain criterion, the instructor may conclude that he is attaining the instructional objectives and that he is doing an effective job. If less than the desired percentage of students attain criterion, then the instructor must conclude that his instruction has not been as effective as he desired it to be, and he must than proceed to decide whether he should change the instruction, the CRM instrument, or his objectives for his next attempt at teaching the material.

In some situations it may be worthwhile to repeat the instruction for all of the students who did not pass, and to continue repeating instruction until all of these students can attain a criterion score.

Figure 1-4 demonstrates a step by step procedure of CRM.

1. State objectives.
2. Prepare CRM instrument to measure objectives.
3. Teach to accomplish objectives.
4. Administer and score CRM instrument.
5. If any student scored above _____%, he has mastered the instruction.
6. If _____% of the students score above _____%, instruction is effective.
7. Decide if a change is needed in objectives, CRM instrument or the instruction.

**Figure 1-4**
**Steps in CRM**

In Figure 1-4, the criterion levels were not specified because it is the decision of each instructor as to what level the class should attain.

Figure 1-5 shows a model of the decision-making process associated with CRM and contrasts it with the process traditionally followed in NRM.

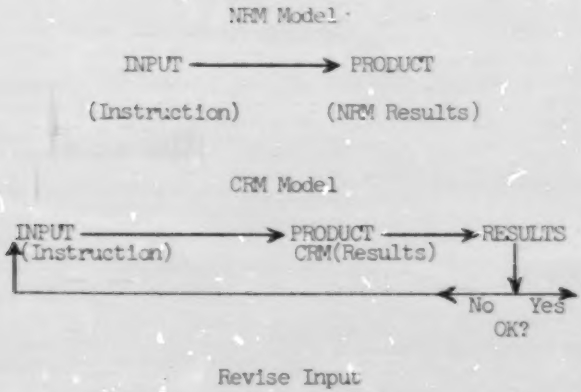


**Figure 1-5**
**Decision Making Process in NRM and CRM**

From Figure 1-5, it may be observed that in NRM there is no attempt made to revise instruction on the basis of the product results as measured by NRM. However, in the CRM process, revisions occur if the test results indicate that the instructional objectives are not being accomplished.

## Differences in the Methods of NRM and CRM

It is difficult for the layman or teaching practitioner to conceive of educational testing as having differing philosophies. The concept of a philosophy of measurement is not easy to contemplate. However, measurement procedures do follow their philosophies and there are striking differences between the philosophy of NRM and that of CRM. Some of the differences are described in the paragraphs below.

**Trait or ability to be measured.** In NRM, the trait or ability to be measured is assumed to be present in varying degrees in different individuals. It is the purpose of NRM to order those individuals on a continuum ranging from highest to lowest in terms of the amount of that trait or ability the learner possesses. In CRM, the trait or ability is assumed to be present in either a sufficient or in an insufficient amount in different individuals. It is the purpose of CRM to separate those individuals who have attained a prescribed level of mastery of the trait or ability from those who have not.

**Range of scores.** In NRM, students test scores range from a low which is approximately equal to the chance level of the test to a high

which is often equal to a score of 100%. Thus if a group of students complete a four-response multiple-choice test, it can be expected that their scores will range from the chance level (25%) to 100%. Each student will have a score somewhere on the continuum from 25% to 100%. Since scores may be ranged on a continuum from chance to a perfect score, NRM scores are often described as being *continuous data*. A desirable characteristic of an NRM test is considered to be a wide range of scores, as measured by a high standard deviation.

CRM scores are considered to be passing if the student attains criterion or above criterion score and are considered to be not passing if the student does not attain a criterion score. CRM scores can take only one of two values. The two values are sometimes specified as pass-not pass, pass-fail, go-no go, yes-no, or adequate-inadequate. The two-value scoring of CRM is frequently referred to as producing *dichotomous data*. However, it could be logically argued that some of the best CRM instruments are those on which everyone receives the *same* score of pass.

**Difficulty of items**. Most test theorists believe that norm referenced test items of medium difficulty will produce the greatest discrimination, the most information, and will contribute most to the test's reliability. This means that for a short answer completion test, the ideal test item would be on to which half of the students respond correctly and the other half respond incorrectly or omit the item.

Neither psychology nor common sense would support asking students a question with advance knowledge that half of them will not obtain the correct response to the question.

Although the actual difficulty level of CRM instrument items depends on the ability of the group of students involved the level of mastery required, and the objectives of the instructor, traditionally CRM items are relatively easy test items. It is not unusual for the 90-90 criterion to be used. The use of this criterion implies that the test is designed so that the items should be easy enough that 90% of the students will be able to score 90% or above on the test, assuming, of course, that instruction was effective.

**Domain of instruction**. Although it is difficult to make any infallible generalization concerning the domain of instruction measured by the two types of tests, it is fairly safe to say that NRM has most often been used for measuring learning of factual information and concepts, usually referred to as the cognitive domain. Although CRM may be used for measurement in the cognitive domain or to measure the acquisition of attitudes and skills (the affective domain), the nature of CRM makes it especially useful for measuring learning in the physical skills and competencies that are included in the psychomotor domain. CRM attempts to measure what a student can do, rather than what he knows.

**Discrimination**. NRM tests attempt to order groups of students from high to low. An NRM test item is considered to be a good item if those who do well on the test do well on that item. Item analysis is a

procedure through which a test constructor looks carefully at each item to determine if the item discriminates between good and poor students. Items that do not have this quality are discarded and do not remain on the test.

CRM can not use conventional item analysis, but there have been attempts to substitute the before and after property of specificity for discrimination. The best CRM items are sometimes considered to be those that students do not answer correctly in a pretest situation before instruction begins but can respond to correctly in a posttest situation after instruction has been completed. Items that are answered correctly only in the posttest situation are said to be specific.

However, item analysis may be applied to CRM in an attempt to find which items discriminate between mastery and non-mastery of each of the instructional objectives.

Reliability. The reliability or precision of measurement is a prime consideration for NRM. Most often, reliability estimates for NRM instruments are obtained indirectly by correlational coefficients, since reliability can not be obtained by more direct methods. Since reliability in CRM is not considered to be an overriding concern, most CRM instruments are constructed without any attention to their reliability. NRM instruments are usually relatively long tests, since the degree of reliability is directly related to test length. Since CRM instruments are not concerned with reliability, they may be shorter tests. Some CRM exercises are essentially one-item tests. Several research papers which propose methods for calculating reliability for CRM exercises have been published in measurement journals.

Validity. There are many methods for determining the validity of an NRM instrument. Perhaps content validity is the most frequent validity determination for NRM achievement tests. Contest validity attempts to demonstrate that the items covered on the test constitute a representative sample of the material convered during instruction. Although some experts propose other validation techniques to be applicable. Curricular validity determination is accomplished by keying certain test items to each of the instructional objectives. This, of course, is the essence of the method used in CRM.

Previously acquired skills. In NRM, students must often use previously acquired skills to respond to items so that they may demonstrate the broad understandings measured by NRM. CRM usually measures only instructional objectives and requires no previously acquired skills.

Comparisons. NRM measures a student's performance in relation to that of the group and also to that of each of the other students. CRM encourages competition with oneself to acquire proficiencies. It attempts to find what the student can or cannot do. The student's score is compared to the criterion.

**Instruction related to the test.** Instructors who teach to a NRM test try to maximize the amount of material covered. Often the objectives of NRM imply that the student is to be provided, by means of a broad survey of the subject, a thorough familiarity with all aspects of the subject. CRM tries to maximize the percentage of students who master the objectives of instruction.

**Score.** The score received in an NRM test is usually the number of items answered correctly or the percentage of correct responses. As previously indicated, the only score a student receives on a CRM test is either of two dichotomous scores, pass or non-pass.

**Function.** NRM measures the amount of knowledge learned by ranking students from high to low. CRM evaluates the effectiveness of instruction.

## Advantages

Many articles have been written in which the authors declare NRM to be immoral and proclaim CRM to be the only humane way to evaluate students. The rationale for these articles appears to be it is an inherent characteristic of NRM that half the students must miss each item and half the students must fall below the median. This approach does not encourage the type of success that enhances motivation or learning. The critics of NRM also fault student-vs.-student competition and consider the competition of a student with himself or with a criterion to be healthier and non-destructive.

Certainly the potential for evaluating instruction is greater in CRM than in NRM, because traditional NRM has never really been concerned with the evaluating of instructional effectiveness. Furthermore, only CRM offers the potential to provide the data for effectively revising instructional content.

The NRM model for measurement has been one which has been preoccupied with aptitude, selection, and prediction. The CRM model is concerned with evaluating and revising instruction. CRM can lead to more meaningful statements than the NRM model when the criteria are obvious and simple.

## Limitations

Defenders of NRM claim that the criticized aspects of NRM are *not* inherent in the NRM model. They allege that the fault does not lie in the model itself, but rather is the result of poor testing practices. For that reason defenders of NRM believe that CRM is a fad that will soon fall back into disuse.

There seems to be some doubt among measurement specialists about the versatility of CRM and about its ability to measure complex behaviors. Indeed, some educators believe that in some types of instruction, there are no identifiable criteria. Many measurement specialists feel that CRM is practical only in those few areas of

achievement which focus on the cultivation of a high degree of skill in the exercise of a limited number of abilities. Others consider CRM to be a measure of the degree of mastery of material taught in a specific time frame *before* the student progresses to higher level.

Some practical disadvantages to CRM are that reporting systems vary and must be interpreted to the parents of children moving into new school districts. Comparisons of performances on CRM between school districts are not yet readily available. Further work is necessary to determine if CRM constitutes a valid measure of academic progress.

An objective evaluation of the idealistic foundations of CRM raises a few perplexing questions. First, can most of the students attain most of the objectives of instruction in a well taught course? Second, will test items based on separate objectives of instruction assess how much a student has learned more accurately than test items requiring the understanding and synthesizing of several objectives?

Furthermore, some educators feel CRM does not tell us everything we need to know about achievement because a single criterion does not allow for any student to attain excellence nor does it tell as much as should be known about deficiencies in achievement.

Although CRM can be used in most educational situations, some test theorists have been quick to point out that NRM is applicable in virtually all teaching situations. Furthermore, since NRM is firmly established in American education, practically all present standardized tests are NRM instruments.

Nevertheless, the attention educators are giving to CRM is continually increasing and more knowledge about the process of CRM may cause educators to view it as a superior evaluative procedure.

## Chapter 2

## ALTERNATIVES TO TESTING: SELF-SCORING TESTS

The self-scoring test, also known as adjunct autoinstruction, has been used in American schools, although not extensively, since the early 1900's. The self-scoring test is a series of questions designed to help determine whether or not a student's learning is progressing satisfactorily.

The questions themselves are prepared in multiple-choice form with the incorrect alternatives selected from common misunderstandings. Through one of a variety of rather ingenious techniques, the student is given immediate knowledge concerning the correctness of his response.

The strategy for the self-scoring test is that students who select a correct response may proceed to the next question. Students who are unable to select the correct response on their first response to an item, must continue to select from the remaining alternatives until the correct response has been given. Feedback messages, usually consisting of letters or symbols, inform the student of the correctness of his response.

The questions in a self-scoring test do not necessarily cover everything in any one lesson or unit and may very well jump back and forth from one point to another. The purpose of these questions is to help the student determine whether or not his learning has progressed satisfactorily.

The self-scoring test is scored by counting the total number of responses required to answer all of the questions correctly. The counting of one point for each response is similar to the scoring procedure used in golf, where one point is counted toward a player's score for each stroke a player makes in his attempts to advance the ball.

The logic of self-scoring testing is supported by numerous principles of educational psychology. Its scoring techniques are readily adaptable to numerous technological or chemical devices. Furthermore, when educational experiments are conducted comparing the learning that occurs as a result of self-scoring testing with traditional testing methods, the results almost invariably indicate advantages for the self-scoring tests.

The self-scoring test technique provides a non-punitive type of testing. It may be utilized as a technique for determining if instruction is proceeding effectively (formative evaluation) or to determine if a student has mastered the unit he is studying (summative evaluation).

Self-scoring tests may be administered frequently to stimulate students' study. The feedback helps the student to differentiate between right and wrong answers and extinguishes misconceptions. It has been said that minute for minute, no form of learning provides more efficient learning than a test. The immediate feedback potential of the self-scoring test enhances the learning potential of the testing process.

It is fair to say that if any learning exists in the testing situation, it probably occurs as a result of the feedback a student receives after the response has been made. In traditional testing the feedback occurs days and sometimes even weeks after the response, if it is provided at all. In self-scoring tests the feedback for each item immediately follows the student's response to that item.

The self-scoring test situation is one in which the test is administered, and simultaneously automatically teaches. If a student's answer is correct, he is informed and reinforced immediately. If the answer is wrong, he carefully considers the remaining alternatives before he makes another response. The question is kept before the student until a correct response is given. He must get the correct answer to each question before he can go on to the next. When he does get a correct answer, he is immediately informed that he has done so.

The questions that are used in self-scoring tests often deal with methods, conclusions, applications and other concepts that more or less involve global understanding of the entire unit being studied. Such questions naturally fit in the multiple-choice format in which the wrong alternatives are common misconceptions or misunderstandings that students could have developed during the course of instruction.

Aside from the learning advantages that self-scoring tests offer, the tests score themselves and reduce clerical work for the teacher. Since the scoring technique increases the range of scores with a resulting greater standard deviation, the reliability of a test administered in a self-scoring format is higher than the reliability of the same test administered in the format of a traditional multiple-choice test. Students taking a self-scoring test must respond again after each correct response. This is similar to increasing the length of a test without really adding more questions. The seemingly greater test length also favorably affects reliability.

The self-scoring test, if administered under supervised conditions in an uncrowded room, is virtually cheat proof.

## Varieties of Self-Scoring Tests

The self-scoring test format can be adapted to a wide variety of technologies. There are literally hundreds of products that may be purchased that employ the self-scoring test technique. Some of the most common are described below.

**Erasure card.** A student is provided with a form which requires him to erase carbon spots on a multiple-choice answer sheet according to which answer he thinks is right. Each spot is partly covered with easily removed carbon shield which, when erased, reveals a letter to indicate the correctness of response. For example, if the student's response is right, he finds an "R" under the carbon spot. If his response is incorrect, an "X" appears.

**Latent image marker.** Multiple-choice answer sheets are treated with a chemical so that when the student marks his response with a felt pen, feedback messages appear. The messages are letters or symbols and are similar to those used on the erasure cards. Short answer completion questions also may be used and the latent images are words that form the correct answer.

**Chemical paper.** Some self-scoring tests are printed on paper treated in such a way that the response choices on the answer sheets change colors when a correct response has been made. Either a multiple-choice or a true-false format may be used. The correct response space turns green when moistened and an incorrect response causes that response to turn red when it is chosen by a student.

**Tab tests.** A perforated tab is removed by the student to reveal a feedback symbol or feedback message.

**Electric grid.** A very simple teaching machine can be made by wiring an electrical grid in such a way that a correct response will produce a completed circuit. When a question card is placed over the grid, the student inserts a pointer into a hole in the card next to the answer. If the pointer is placed in the hole next to the correct answer, the grid activates a buzzer or a light.

**Pressey punchboard.** This device is similar to the punchboards that are used in chance games. The Pressey punchboard requires a student to push a pencil or key into a hole corresponding to his choice of an answer. If the response is correct, the pencil breaks through a paper cover sheet. If he is wrong, the pencil merely marks the cover sheet.

**Microfilm teaching machines.** A microfilm reader is equipped so that it can be programmed to display questions and feedback messages and advance according to instructions coded on the microfilm. Frequently these devices are equipped with score-keeping equipment or response recorders.

**Computer-assisted instruction.** The ultimate in teaching machines is computer-assisted instruction. The computer is programmed to cause a teletypewriter or cathode ray tube monitor to display questions and feedback messages. Such sophisticated hardware offers complex branching capability, a complex record keeping and analysis system, and can be equipped with audio and visual display accessories.

**Other devices.** Self-scoring tests are sometimes administered by means of filmstrips or tape recordings. The selection of a response causes the filmstrip or tape to advance to an appropriate feedback message. Selection of the correct response cues presentation of the next question.

**Summary.** The self-scoring test assists students in learning and motivates them to learn by providing instant feedback. The student is invited to reread the questions and correct his errors while he works. The completed test gives the instructor a clear record of the student's problems and strengths. The self-scoring test format provides a very reliable measuring technique and frees the instructor from the mechanical chore of checking each answer on every student's paper.

Self-scoring tests are readily available and the use of them is supported by a plethora of convincing research studies. It is rather surprising that such a simple and effective device has not been overwhelmingly adopted by educators.

Perhaps the reluctance of educators to accept the self-scoring test has been due to the cultural inertia that makes them reluctant to try any new teaching or testing technique.

It is fair to say, however, that a teacher's initial attempts to use self-scoring tests may cause the students to experience somewhat greater test anxiety than they experience in traditional testing. A confirmed principle of testing is that anxiety increases when an unfamiliar test form is introduced.

However, with repeated use, the worth of self-scoring tests as reliable measuring devices and as tools for producing effective learning is one of the few certainties in education.

## Chapter 3

## ALTERNATIVES TO TESTING: THE CONTRACT METHOD

The contract approach to evaluation has been used with classes of students from nursery school to graduate school. Although the contract is a popular method of evaluation in schools everywhere today, there is surprisingly little published information about it.

The contract method of evaluation is a very simple procedure. At the beginning of the marking period, the teacher and the student meet in what might be described as a conference or interview. During this meeting, the teacher and student jointly decide what work the student will do in order to qualify for whatever grade the student chooses to work for. The quantity of work a student must complete to receive an "A" grade will necessarily be greater than the quantity of work that would be required for a "B" grade, etc. The conditions of the evaluation are clearly stated in a written contract.

Although the student is guaranteed that he will receive the requested grade if the required work is completed satisfactorily, contracts usually contain a statement that if the work is not completed satisfactorily, the student will receive the grade that the instructor feels is appropriate.

Contracts are signed at the beginning of the grading period and are evaluated at the end of the grading period. Contracts may be changed by mutual consent of the teacher and the student at any time during the marking period.

Contracts may be written to apply to whole classes. However, sometimes individual contracts are designed by each student and agreed to by the teacher.

Contracts for whole classes. When the contract is applied to the entire class, the contract is written so that it specifies the quantity of work that is required for each letter grade.

The criteria listed to be evaluated may include such tasks as projects, reports, research papers, community service, books to be read, as well as attendance and even citizenship.

The contract may be written by the teacher and presented to the students or may be worked out as a group decision by the entire class.

Figure 3-1 is a typical whole class contract for a college class. It will be noted that less attendance is required for an "A" grade than for other letter grades. This is because of the great amount of time that "A" students will need to complete all of their contractual tasks.

SOCIAL STUDIES 102
WORLD HISTORY

| Contract Items (Those marked X required) | Letter Grade A | B | C | D |
|---|---|---|---|---|
| 1. Attend class regularly | | X | X | X |
| 2. Read and report on one book dealing with ancient history | X | | | |
| 3. Read and report on one book dealing with the Dark Ages | X | | | |
| 4. Read and report on one book dealing with the Industrial Revolution | X | | | |
| 5. Write a 5 page report on serfdom | X | | | |
| 6. Write a 5 page report on the Protestant Reformation | X | | | |
| 7. Write a 5 page report on the Roman Empire | X | | | |
| 8. Write a 5 page report on Ancient Greece | X | | | |
| 9. Make a chart contrasting socialism and communism | X | X | X | X |
| 10. Make a chart comparing the conquests of Alexander the Great, Julius Caesar and Genghis Khan | X | X | X | |
| 11. Construct a time line chart showing important periods and historical developments from 1100 to 1900 | X | X | X | X |

I wish to contract for the grade of _____. If any of the requirements listed above are incomplete or unacceptable, I will accept the grade the teacher feels is appropriate.

Student _____

Teacher _____

Date _____

**Figure 3-1**
**A Whole Class Evaluation Contract**

The lack of an attendance requirement for "A" students is characteristic of an honors class and writing the contract with no attendance requirement for the "A" students will give the class the atmosphere of an honors seminar.

From Figure 3-1, it can also be noted that generally a higher grade requires more tasks to be performed. A possible exception to this generalization is sometimes found in the differences between the "C" and "D" grades. Often the requirements for the "D" grade are as great as the requirements for the "C" grade because there is a general tendency to try to discourage students from contracting for a low grade.

Semi-contracts. When a contract is applied to a whole class, the emphasis is on the quantity of work, rather than the quality of work. Instructors who wish to use the contract method of grading, but who also require some degree of mastery from their students may resort to the semi-contract. The semi-contract is similar to the contract in most respects, except that one condition of the contract is the requirement that the student must achieve at least a certain minimum score on an examination in order to receive the contracted grade. For example, a contract might be written to also require a score of 80% on the final examination for a grade of "A."

Contracts used for individual students. An alternative approach to writing one contract for the entire class is to allow each student to write a contract and to present it to the teacher. The teacher may either sign the contract in its original form or make suggestions for changing the contract.

When each student designs his own contract, each contract calls for a different variety of activities to be completed by each student. Each contract contains an agreement of how grades are to be determined.

Figure 3-2 shows a contract that has been written by a student for evaluation purposes. The tasks selected by the student are tasks that no other student would probably select and are relevant tasks for the student. The teacher and student work together to reach an agreement on the criteria to be used for evaluation and whatever levels of proficiency are required.

Contracts may be and usually are evaluated by the teacher. However, the contract may be evaluated by the student, by the feedback the teacher solicits and receives from the class, or by an outside source. For example, if the class is studying a unit in government and one of the requirements of the contract is to assist a community official, the terms of the contract may be evaluated by the community official who is being assisted.

Advantages of contracts. The contract method of grading eliminates the "destructive" competition that matches student against student in a battle for class rank. The student evaluated by the contract method is not competing against other students, but rather is competing against

Student Learning Contract

Student's Name _____ Beginning Date _____

Unit to be Studied _____

Expected Completion Date _____

I. What objectives do you expect to accomplish by studying this unit?

II. What activities do you intend to perform to accomplish these objectives?

III. What evidence do you intend to produce (term papers, reports, outlines, etc.) to demonstrate that you have accomplished your objectives?

IV. Upon satisfactory completion of the work listed above, a grade of _____ will be awarded.

Signature of Student _____

Signature of Teacher _____

Signature of Parents _____

**Figure 3-2**
**A Contract for an Individual Student**

himself and against his contract.

Since the student knows what is expected of him and since there are rarely any tests in the contract method, the anxiety associated with testing and with uncertain evaluations is eliminated.

It can also be argued that since the quantity of work performed is the principal criterion used in the contract method, some of the subjectivity that is a side effect of some other forms of evaluation is eliminated.

An argument for the contract method of grading that is essentially educationally sound is that this type of evaluation encourages diversity in the tasks students are to perform. The tasks will hopefully be tasks that are relevant to the student. The student sets his own goals and follows his own learning patterns.

Disadvantages of contracts. Perhaps the principal philosophical disadvantage of the contract method is that the evaluation is based on the quantity of work performed and little consideration is given to the quality of work. It is difficult to require challenging quantities of work without giving assignments that amount to busy work. Likewise, it is difficult to find creative ways to measure the quality of the work in the context of the grading contract.

When an instructor attempts to design a contract for the whole class, it will probably be very difficult for him to determine an appropriate quantity of work for the various letter grades. Often, the requirements are set too low. This results in the teacher being forced to award high grades to all his students. If other teachers are not awarding such high grades, the failure to set quantitative standards high enough can result in poor staff relationships. Conversely, setting quantitative standards too high can cause the student to be overworked, bored, or inundated with busy work.

When attempts are made to introduce quality controls into the contract system, it is often necessary to go back to many of the subjective judgments that the contract method seeks to avoid.

The contract method of grading encourages neither quality nor excellence. Teachers who use the contract method are frequently surprised and frustrated when they find that some of the students who appear to be the brightest and some of the students who contribute most to the class are the ones who have contracted for low grades in order to escape some of the busy work that is required for higher grades.

An evaluation schedule. A variation of the contract that may be used by an enterprising instructor is the evaluation schedule. This is essentially the same as a contract, but contains quality controls since each of the tasks is evaluated by the instructor according to his judgment of how it was performed. Although this system may be more reliable and consequently more fair than the contract, it requires a

great number of subjective judgments to be made by the teacher.

Figure 3-3 is an evaluation schedule that has been used in an introductory class in educational psychology. It will be noted that the class is one which places an emphasis on the performance of tasks, but also emphasizes the written evaluations of tests and quizzes. The total point score is the criterion for grading. If a student receives more than 550 points, that student will receive an "A" grade for the course, etc. Students who score lower than 250 points do not pass the course.

It can be readily observed from Figure 3-3 that although the evaluation schedule has some of the features of a contract, the evaluation schedule requires that a large number of subjective judgments be made concerning a student's performance rather than just one.

**Summary**. The contract method of grading is growing rapidly. Sometimes contracts are written by the teacher to apply to the whole class and in other cases contracts are prepared by each student for individualized evaluation.

Contracts promote a constructive form of competition, reduce subjectivity in grading, and cause students to be less anxious about their grades. Since contracts evaluate on the basis of quantity rather than quality, it is difficult to avoid assigning "busy work" as a part of the contract. It is also difficult to determine how much work should be assigned to provide a fair, yet challenging, student work load.

Evaluation schedules have many of the characteristics of contracts, but require the instructor to evaluate many criteria subjectively.

27/2

Educational Psychology 202

Evaluation Schedule

| | Outstanding(++) | Good(+) | Poor(✓) | Unsatisfactory(-) | |
|---|---|---|---|---|---|
| 1. Attendance and Class Participation | 50 | 40 | 20 | -50 | ____ |
| 2. Tutoring or Educational Service ( ) | 50 | 40 | 20 | -50 | ____ |
| 3. Lab School Participation | 50 | 40 | (30) 20 | -50 | ____ |
| 4. Oral Book Discussion ( ) | 25 | 20 | 10 | 0 | ____ |
| 5. Colloquia ( ) | 25 | 20 | 5 | 0 | ____ |
| 6. Seminar Group ( ) | 25 | 20 | 5 | 0 | ____ |
| 7. Written Assignments | 50 | 40 | 10 | 0 | ____ |
| 8. Total Quiz Score | | | | | ____ |
| 9. First Test (50 items, 100 points) | | | Score ____ | 550 | A |
| | | | | 500 | B+ |
| Second Test (50 items, 100 points) | | | Score ____ | 450 | B |
| | | | | 400 | C+ |
| Final Test (75 items, 150 points) | | | Score ____ | 350 | C |
| | | | | 300 | D+ |
| | | | Total Test Score ____ | 250 | D |
| | | | | Below 250 | Nice Try |
| | | | Grand Total ____ | | |

I agree to complete the assignments as indicated above. Having completed the assignments satisfactorily, I will receive the grade according to the above contract. Contracts are changed by mutual consent.

_____ _____(Student) _____ (Professor)

_____ (Date)

1. Class attendance is required, expected, and should approach 100%. Quality of class participation is also important.

2. Students may elect an instructional or educational service project. This involves working with live students for approximately ten one-hour sessions. A log should be kept which includes objectives, progress, and evaluation of your goals. The final report will indicate what your goals were, how they were carried out, and an evaluation of your experience.

3. Each student actively participates and assists in instructing a group of students at the laboratory school. A written report includes a description of where and how you helped and an evaluation of what you have observed in the class.

4. Present a five minute oral report of a book or other publication. The report should include a brief summary of the book which should be no longer than five minutes. Then, ask the class *challenging confrontation type questions* concerning the book. Visual aids will be helpful and are recommended.

5. Attend and report on three education or psychology colloquia.

6. Attend a class seminar at assigned times. Seminars will number 3 and will last for about 60 minutes each.

7. Papers are evaluated on the basis of effort and communications skills.

**Figure 3-3**
**An Evaluation Schedule**

## Chapter 4

## ALTERNATIVES TO GRADING

Of all of the alternative measures to evaluation, the opposition to grading and report cards has been the most emotional. Several individuals, some of them educators, have developed a national reputation and have made quite a bit of money by presenting emotional arguments in paperback books or in journal articles. The polemics they present are not always completely logical but the presentations claim that grades are either immoral or irrelevant or both. A national movement, which is gathering strength, would have grades and report cards eliminated from all American schools.

In the past, grading has been assumed to be a function that somehow promoted growth. In the next few pages, an attempt will be made to present an outline of some of the arguments for and against grades and to describe some of the alternative procedures that may be used if letter grades and report cards are eliminated. The majority of U.S. schools still use letter grades. This is true in the primary grades, elementary schools, junior high schools, high schools, colleges, and universities. Recent surveys have demonstrated that 62% of the students of junior high school age and a majority of the parents of students prefer letter grades over any of the other forms of evaluation.

The final part of this section will consider alternatives to grading. Letter grading is the established pattern for which alternatives will be proposed. However, such marking systems which use symbols to convey achievement in terms of *O*utstanding, *S*atisfactory, and *U*nsatisfactory, or *E*xcellent, *G*ood, *P*oor, or *C*ommendable Progress, *S*atisfactory Progress, and *M*inimal Progress are not innovative and are really not very different from traditional letter grading. Therefore, these marking systems will not be considered here.

The logic of the reasons for and against grading is presented in what is intended to be an objective, logical, unemotional manner. Furthermore, the descriptions that follow are intended to relate to the context of practical situations, rather than to the idealistic situations that are often described by critics of contemporary education and opponents of current evaluation practices.

### Arguments For and Against Grading

Whoever attempts to explain the theoretical positions that support grading from a psychological point of view is immediately in trouble. Although certain bases have been proposed as positions supporting grading and report cards, these bases themselves are controversial. Although they unquestionably apply to small animals performing physical learning tasks in an animal laboratory, most of the principles that constitute these bases have not been demonstrated with humans involved

in verbal learning tasks in classroom situations.

Likewise, it has always been difficult to prove or even demonstrate that grades are in any way psychologically damaging to students.

**Arguments for grading**. Probably the principal justification for grading students is the notion that the "reward" of a grade is an incentive that enhances learning and promotes study. This notion is somewhat supported by principles of operant conditioning, behaviorism, and reward psychology. The reinforcement that follows a certain type of behavior increases the probability that the behavior will occur again. The theoretical rationale for reinforcement requires that the reinforcement requires that the reinforcement should follow the behavior immediately. Of course, grades are not given immediately after the responses they are supposed to reinforce. However, although reinforcement theory is fairly well developed, there is not complete agreement that reinforcement has to be immediate to be effective.

A further justification for grades is based on studies of motivation. Although working for a grade may have nothing to do with the subject that is being studied, providing students with tangible or intangible incentives for learning has been demonstrated to be effective in producing learning in psychological experiments. The process of motivating learning through reward is called extrinsic motivation.

Although most psychologists believe that learning may be enhanced by any type of motivation that is effective, some critics of grades believe that motivation for learning should be intrinsic, which means that motivation results from a desire to learn more about the subject matter.

It is idealistic to assume that all children can be motivated to learn simply as a result of their thirst for more knowledge. However, it is equally naive to believe that all students, particularly the low achievers, can be motivated by the slim prospect of the reward of a high grade.

Another argument for grading students comes not from psychology, but rather from economics. Several economic theories are based on what is known as an economic system of social justice. Economic rewards to citizens are often based on some type of system. In the case of grades, the system works in such a way that the students who work hardest and achieve most are supposed to receive better rewards. Advocates of grading consider it to be a system which promotes justice in an economic sense. Presumably, groups of people are somehow supposed to be more content when they feel that they are working in a system where rewards are distributed in a just and equitable manner. It is possible to think of the grading system as a micro-economic model in which benefits accrue to those who produce the most. Such a system is said to be not unlike those systems the student will encounter in the world of work. However, it is fair to say that several studies, particularly those conducted with college students, have demonstrated a fairly low

correlation between grades and later economic success.

An argument for grading that is hard to dispute is that letter grades are an efficient means of communication. Although there are probably as many methods for assigning grades to students as there are teachers, the meaning of letter grades is universal. Everyone knows that an "A" is a good grade and everyone knows that a "D" indicates poor work. Although grades are efficient in that they convey information succinctly, grades in themselves do not in any way indicate exactly what it is that is causing the student trouble nor do they indicate what it is that a student is accomplishing well. However, grades do provide a simple system that facilitates reporting a student's achievement to future employers and college admission officials, as well as to parents, students, and teachers.

An advantage of grading that is practical, although not very humanistic, is its administrative convenience. Grades are a convenient tool for crucial decisions regarding college admissions, participation in extra-curricular activities, financial aid, ability grouping, and various academic honors.

Other arguments for grading are that grades enhance discipline, prepare students for competition in the world of work, and are a fairly reliable index of academic achievement. However, the most convincing argument for grades comes from school officials and teachers who consider that grading is achieving its purpose in schools. Many school administrators find there are no complaints with grading by either students or parents and consider grading to be the best system for student evaluation. As has been mentioned, the majority of students and parents surveyed indicate that they prefer letter grades as a method of evaluation. Many schools that had once abandoned grading have gone back to it because they feel that they have learned that it constitutes the best available system for student evaluation.

Arguments against grading. The arguments against grading are just as practical and as convincing as the arguments for grading. However, these arguments are usually presented in a somewhat more emotional context and are based on ideas developed by humanistic or existential educators.

The most pervasive argument against grading is that many students cannot possibly get good grades even with maximum effort. Consequently they continually receive low grades, which is degrading to them. A categorizing of students is said to develop from grading procedures and thus grading serves to label low achievers as inferior persons. The process of grading lowers the self concept of low achieving students to the benefit of their more intelligent peers. It is argued that grading is an ego-damaging process that brands children as failures and generates rebellion.

The reliability and accuracy of grading are questionable, since grades represent subjective judgments of student achievement by the

teacher. It is said that grades are partly determined by such variables as effort, citizenship, attitude, handwriting skill, race, sex, religion, background, appearance, and dress of the student.

Grades are considered by many of their opponents to promote dishonesty and encourage cheating. The type of verbal learning that grades encourage is said to be mostly memorization and learning of factual materials. The student, it is said, only learns the materials that he feels will be covered by the test. Furthermore, grades are considered to cause the teacher to emphasize the type of learning which is measurable and to deemphasize teaching of attitudes and concepts that can not be easily measured by paper and pencil tests.

From a psychological point of view, grades are said to increase pressure and anxiety on the student, although some psychological experiments have demonstrated that learning occurs best when students experience a mild amount of pressure and anxiety.

Although the grade is not the real value in the course, it often is considered to be. Such an attitude causes students to take courses and to enroll in classes where they will receive good grades, rather than to take courses that will benefit them. Also, the teacher is placed in a dual role of teacher and evaluator, which amounts to having the teacher function as the student's critic as well as his helper.

It is often argued that grades are irrelevant to the learning process and that studying for tests restricts the creative endeavors of students.

Since grades mean different things to different instructors, their interpretation or the information they convey is subject to many factors. It is doubtful if a single mark can convey to parents and students the necessary information that is needed to enhance learning. The excellence a student may achieve in one unit is frequently masked by the averaging process. While a student may do outstanding work in one area, his performance in that unit may be lost sight of because it was obliterated when it was averaged with his lower performance in another unit.

## Emphasizing Strengths and Eliminating Weaknesses

There is generally, although not always, a consensus among both opponents and advocates of grading that some sort of evaluation of the learning process is very desirable. Obviously the type of evaluative process that most teachers, principals, and parents would approve of is one which would emphasize the strengths of grading and eliminate some of its weaknesses.

It is probably correct to say that when some of the weaknesses of the grading process are eliminated by the introduction of a new system of evaluation, the new system will have weaknesses not apparent in the old system.

The alternatives to letter grading will be described in the next few pages. If the reader is contemplating abandoning letter grades or is thinking of introducing a new system of evaluation, he would do well to consider the effects that the new system will have on the learning process and carefully evaluate the new system in the light of his educational philosophy.

In spite of the emotional appeals made by the opponents of grades and the steadfast reliance on letter grades by advocates of that form of evaluation, probably grades are neither the ogres or the good fairies that either of these extremists would have us believe they are. Nevertheless, it is advantageous for all educators to be informed of alternative methods of student evaluation and to carefully consider changes in evaluation from the points of view of the objectives of instruction and the educator's personal philosophy.

## Alternative Methods

**The parent-teacher conference.** The face-to-face meeting of the parent and the teacher, which is utilized in many elementary schools today, has been described as the ideal means of reporting student achievement. In this method, it is possible for parent and teacher to discuss openly the problems and progress of student achievement in each subject matter area.

The conference has been, and continues to be, the fastest growing procedure for the reporting of student achievement in American schools. It is well established in kindergarten and primary grades, and continues to spread to elementary and high school.

In addition to the fact that the parent-teacher conference provides an opportunity for an open-ended detailed report of all aspects of student achievement, it also sets up a two-way communication between school and home.

The disadvantages of the parent-teacher conference approach to evaluation is the manner in which the conference is usually conducted. Most parents attend expecting to listen, and spend most of their time doing so. Most teachers who schedule conferences expect to spend most of the conference time talking, and likewise do so. However, the chief disadvantages of the conferences are in the time they require and the difficulties in scheduling them. While the conference method may be practical for an elementary teacher who has 25 students in her class, it is not an advisable alternative for the secondary teacher of 200 students. Furthermore, since conferences are scheduled during the school day and require the attendance of the teacher, several school days which might have been used for instruction are devoted exclusively to conferences between parent and teachers. In addition to the loss of instruction time, the difficulties for working parents in obtaining child care may cause ill will between parents and educators.

**Narrative reports and letters to parents.** Some school systems have

found that narrative reports are better for their evaluation purposes than are report cards. A narrative report may take the form of a letter to the parents or may be merely a form in which the evaluation is written in prose, rather than letter grades or check marks.

The narrative report has been utilized by schools since 1933, when the schools of Newton, Massachusetts, abolished all report cards and replaced them with individual letters to parents.

The narrative report tells what students have achieved and attempts to communicate just how the school and the student's home can work together. On a narrative report, all of the letters of the alphabet are used to detail a pupil's achievement and to point out problems the student is experiencing in different subject areas, as well as in attitudes and work habits.

Symbols and checkmarks are strictly avoided in a narrative report. A letter grade is considered to be meaningless, since it does not tell precisely what strengths and weaknesses are affecting the student's performance. Although the letter is individualized the comments tend to be rather general.

Sometimes narrative reporting utilizes a format that requires teachers to follow a form. The form lists specific areas to be considered in the evaluation. This format requires more specific comments and is probably more comprehensive than a letter but it will result in a report that lacks continuity.

Some school systems have attempted to reduce the time and effort required of teachers in the narrative reporting process. To accomplish this, data processing techniques and computers are utilized. In this procedure, the teacher may either select appropriate comments on a check list or select appropriate comments from a bank of available statements. A computer is programmed to arrange these comments into a meaningful report or letter to each student's parents.

First, the teacher would select appropriate comments from the comment bank. Then the comments would be entered into a computer and a letter to the parents would be prepared by the computer. The letter might look like the one in figure 4-1.

Some interesting variations in the narrative report are possible. Some school systems have space equal to that used for comments by the teacher available for replies from the parent to the teacher. At least one school system sends a brief announcement in the form of a telegram titled "Congrats-O-Gram" to the parents of students doing exceptional work.

The principal advantage of the narrative report over the report card is that it provides much more meaningful feedback to parents, although it is usually not as informative as the teacher-parent conference. However, it is possible to use narrative reports without

Dear Mr. and Mrs. *Johnson:*

We are sending you this letter to indicate to you the progress of your *daughter, Mary,* in her work at our school. The following comments relate to discipline and citizenship.

*Mary -*

*is well adjusted socially*
*is dependable*
*attends school regularly*
*is mature and dependable*

The next statements apply to the study habits and academic skill level of *Mary* in *English.*

*Mary -*

*does assigned work promptly*
*needs to work on punctuation and grammar*
*writes acceptable essays*

We will be pleased to discuss any of these comments with you at your convenience. Please call for an appointment.

Sincerely,

_____

**Figure 4-1**
**A Computer Printout Narrative Report**

the inconvenience of loss of school days that parent-teacher conferences require. The individual attention that students receive through narrative reporting often allows teachers to evaluate strengths and weaknesses of their students more carefully and more comprehensively than they would if grades were merely marked on a report card. The communication between parents and school that the narrative report supplies can be a valuable tool for producing better community relations.

However, narrative reports tend to be more subjective than the grades a student receives on a report card. The grades represent the average of scores on objective tests, but narrative reports represent the judgment of the student by the teacher. Furthermore, not all teachers are capable of writing meaningful evaluations that will be helpful to students. Teachers who are not suited to report writing may find the computer to be of valuable assistance.

There are also some administrative inconveniences that are involved in narrative reporting. Writing good narrative reports takes time and skill. Teachers with a large number of students, such as those who teach in high schools, may find the time required for narrative reports to be prohibitive. School systems may find that the expenditure of teachers' salaries for report writing is more costly than it is worth. Also, it is obvious that narrative reports will be more difficult to interpret than letter grades by anyone who has reason to examine the student's school record.

The advantages of narrative reporting are often quickly recognized by parents. In studies where a survey has been conducted in school systems that have tried narrative reporting, a typical finding is that three-quarters of the parents prefer narrative reporting to report cards.

## Chapter 5

## ALTERNATIVE METHODS OF EVALUATION

A common alternative to the letter grade method of reporting student achievement is the method employed in checklists and rating scales. Instead of the broad categories of learning that are described by letter grades, the attitudes and work habits in each subject matter area are broken down into individual components.

### Checklists and Rating Scales

In the case of a checklist, the individual components are checked if the quality being evaluated is present. In rating scales, the quality is rated on a scale that classifies the quality according to its degree of adequacy.

An example of part of a checklist is shown in Figure 5-1. From the checklist it may be noted that the basic characteristics of the mastery of English, spelling, and handwriting have been listed. If the student has mastered these characteristics, a check will be recorded in the appropriate space. If the student is not proficient, the space for the mark will be left blank. There is a space in each section of the report for remarks by the teacher.

In a rating scale, the qualities are evaluated, rather than merely being checked. The student rating scale in Figure 5-2 uses ratings rather than check marks to indicate the degree to which a quality is present. In this rating scale, the qualities are evaluated on a scale from "needs to improve" to "outstanding." In addition, the student's total capability and his achievement relative to his capability are evaluated concurrently and the results of these evaluations appear along with the ratings of the components.

It is also possible to use a rating scale which combines a teacher's evaluation with a student's self-evaluation, as demonstrated in Figure 5-3.

The rating scale may be superior to the report card because it breaks down the student's achievement into components that are identifiable and understandable. The student is able to discern immediately the strengths and weaknesses of his performance. There are implications of what is needed for improvement.

The checklist and rating scale are not particularly adaptable to computing grade point averages, so it is somewhat difficult to translate the information they contain into information that can be used for college admissions decisions.

Busseron Community Schools
Evaluation Checklist

Student's Name _____

ENGLISH

Teacher _____

____ Uses correct forms in speaking
____ Uses correct forms in writing
____ Masters the mechanics of writing
____ Uses appropriate idiom and diction
____ Expresses ideas effectively in written work
____ Expresses ideas effectively in oral work
____ Knows and uses various sentence patterns
____ Knows the criteria for evaluating audio-visual communication
____ Organizes information and uses it as a basis for his own writing
____ Participates in classroom discussions
____ Fulfills required assignments

Remarks: _____

_____

SPELLING

Teacher _____

____ Passes weekly spelling tests
____ Transfers correct spelling to other work
____ Recognizes relationship of various word forms
____ Can apply spelling patterns to new words
____ Fulfills required assignments

Remarks: _____

_____

HANDWRITING

Teacher _____

____ Masters letter formation
____ Develops rhythm and adequate speed in writing
____ Transfers good writing to other work
____ Fulfills assignments

Remarks: _____

_____

**Figure 5-1**
**An Evaluation Checklist**

SHAKAMAK HIGH SCHOOL
Student Rating Sheet

Pupil's Name _____

| LEARNING AREA | Needs to improve | Improving | Works well | Outstanding |
|---|---|---|---|---|
| **LANGUAGE ARTS** | | | | |
| **READING** | | | | |
| Comprehension | | | | |
| Vocabulary | | | | |
| Interest | | | | |
| Word attack skills | | | | |
| Independent reading | | | | |
| Fluency and expression (reading aloud) | | | | |
| **ORAL EXPRESSION** | | | | |
| Communication of ideas | | | | |
| Creativity | | | | |
| English usage | | | | |
| Participation | | | | |
| **WRITTEN EXPRESSION** | | | | |
| Communication of ideas | | | | |
| Organization | | | | |
| English usage | | | | |
| Creativity | | | | |
| Spelling competence | | | | |
| Handwriting | | | | |
| **LISTENING** | | | | |
| Comprehension | | | | |
| Attentive listening | | | | |
| Critical listening | | | | |
| **MATHEMATICS** | | | | |
| Number facts | | | | |
| Reasoning | | | | |
| Computation | | | | |
| Concept comprehension | | | | |
| Application | | | | |

LOW _____ HIGH
Capability at this time

LOW _____ HIGH
Achievement relative to capability

LOW _____ HIGH
Capability at this time

LOW _____ HIGH
Achievement relative to capability

Figure 5-2
A Student Rating Scale

William Tell Senior High School
Teacher-Student Rating Scale

STUDENT PROGRESS AND EVALUATION

Name of student _____ Subject *Economics*

Instructor *Robert Wilseck*

Goals set by student and teacher: Date _____

Student Self-Evaluation: (#1 = best; #8 = poorest)

| | Student's Opinion | Teacher's Opinion |
|---|---|---|
| I give my best effort to the work I have chosen for this course | 1 2 3 4 5 6 7 8 | 1 2 3 4 5 6 7 8 |
| I spend my unscheduled time wisely and efficiently for the work I have chosen for this course | 1 2 3 4 5 6 7 8 | 1 2 3 4 5 6 7 8 |
| I use my scheduled time for this course to take advantage of the teacher's help | 1 2 3 4 5 6 7 8 | 1 2 3 4 5 6 7 8 |
| I have made an effort to arrive at goals that are important to me | 1 2 3 4 5 6 7 8 | 1 2 3 4 5 6 7 8 |
| I am satisfied with my achievement toward the goals I have set for myself | 1 2 3 4 5 6 7 8 | 1 2 3 4 5 6 7 8 |
| I rate the depth and extent of my independent work | 1 2 3 4 5 6 7 8 | 1 2 3 4 5 6 7 8 |

**Figure 5-3**
**A Student-Teacher Rating Scale**

The checklist and rating scale are strong in relevance and validity since they do communicate directly with what it is that the teacher teaches and what it is that the teacher evaluates.

## Self-Evaluation and Self-Grading

Self-evaluation and self-grading are two distinctly different processes. Self-evaluation allows students to evaluate their own progress either in writing or by means of a teacher-student conference. In the process of self-grading, the student determines his own grade.

It is merely for convenience of grouping topics that both of these processes are included in a single chapter. The two processes are supported by different rationales and have little in common. Although it is possible to use self-evaluation and self-grading concurrently, ordinarily only one or the other of these will be utilized in an evaluation.

## Self-Evaluation

Self-evaluation has been described as a means of student assessment that is of the children, by the children and for the children. Self-evaluation allows the student to evaluate his own progress, either in writing or in a conference with the teacher. The teacher communicates comments and reaction to the self-evaluation. Ordinarily there are no grades in self-evaluation, but some teachers may wish to use self-evaluation as part of the basis for assigning grades. Sometimes it is used in combination with peer evaluation and teacher evaluation, but usually when self-evaluation is used it stands alone as the process by which student performance is evaluated.

In many cases, self-evaluation is open-ended and no structure is involved in the evaluation process. The student's comments are made extemporaneously in writing or in an interview with the teacher.

Sometimes self-evaluation is accomplished through the use of a special form. The form provides guidelines for the evaluation and assures the teacher that certain criteria will be evaluated.

From Figure 5-4, it can be observed that the self-evaluation form is simply a technique that will enable the student to respond to certain evaluation criteria that the teacher considers to be important.

Likewise, sometimes a self-evaluation checklist is used by students. This type of self-evaluation instrument is particularly helpful when several criteria are considered in the total evaluation. A self-evaluation checklist displayed as one column of Figure 5-3.

Advantages of self-evaluation. Through self-evaluation, the student is allowed to evaluate his own progress toward goals that he has set for himself. The student is able to establish his own criteria for the evaluation of his work.

McDaid High School
Self-Evaluation Form

My name is _____ and I am completing this report for the _____ semester.

During this semester, I completed _____ oral reports. I feel that my performance on these reports was _____. To improve my performance on oral reports, I need to _____

_____

_____.

I also wrote _____ themes. I feel that my performance on these themes was _____. To write better themes, I need to _____

_____

_____.

Over all, I feel that my strengths in studying the subject are

_____

_____

_____. On the other hand, I need to make improvements in _____

_____

_____

Teacher's Comments _____

_____

Parent's Comments _____

_____

**Figure 5-4**
**A Self-Evaluation Form**

Promoting self-evaluation, which may be an important learning process in itself, allows a student more freedom of instruction and more responsibility in the evaluation process.

Disadvantages of self-evaluation. Depending on the frame of reference of the teacher, some educators might consider allowing students to establish the criteria for their own evaluation to be a weakness, rather than a strength, in the evaluation process.

Self-evaluation is considered by students to be an improvement over conventional grading procedures until the novelty effect wears off, and after that it becomes less desirable to them.

Furthermore, students frequently become wise to the system of self-evaluation and evaluate themselves higher than they really deserve. This increases the error in the evaluation system and makes subsequent teacher evaluation difficult.

The potential of self-evaluation depends on the way teachers use it. If it is used in such a way as to cause students to reflect on their academic progress, it can be of great benefit to them. If self-evaluation is used as a cop-out to avoid the responsibility of assigning grades, then it probably will contribute nothing to any student's education.

If students are under enormous pressure to achieve, the process of self-evaluation is extremely difficult because, under those conditions, it is hard for them to be objective when they have an opportunity to evaluate their own progress.

## Self-Grading

Self-grading allows students to determine their own grades. Self-grading is frequently used in colleges and universities by teachers who are either ultra-permissive or who intensely dislike giving grades to students.

On the last day of the semester, the professor will tell all of the students to assign themselves the letter grade that they think they should receive for the course. Sometimes they are asked to state in a short essay just why they think the grade they assigned to themselves is appropriate.

Aside from the advantage of eliminating student anxiety about grades, few defenses can be offered for self-grading.

Students consider self-grading to be a cop-out by the teacher since they consider that assigning grades to students is the responsibility of the teacher. Students who are wise to the self-grading system find that it is advantageous to always assign themselves the highest possible grade since they reason that a teacher who was too timid to assign them a grade will undoubtedly be too timid to change a

grade a student assigned to himself.

Self-grading is a method of evaluation that is usually disliked intensely by conscientious students. The inconsistencies of students in the use of the self-evaluation process make it one of the most unreliable of evaluation systems.

## Pass-Fail

For 100 years, some colleges and universities have experimented with the pass-fail system of student evaluation. Other systems which fit in this category of grading are pass-no pass, credit-no credit, and satisfactory-unsatisfactory.

If a student achieves satisfactorily in a course, he is given a grade equivalent to "pass" and his credit in the course appears on his transcript. If his work is not satisfactory, there is no penalty for his having attempted the course. At any rate, neither passing nor failing grades contribute to a student's grade point average. Students have an opportunity to redo failing work.

The rationale for pass-fail evaluation is that it provides incentives for students to take courses that interest them without being concerned about class rank should they fail in the course. The pass-fail system encourages students to try difficult courses that they otherwise would avoid. Also, pass-fail is said to eliminate anxiety among poor achievers by helping them concentrate on what they learn rather than on the grade they earn in the course.

The pass-fail method has been very popular in recent years in colleges, universities, and to some extent in high schools. The results of pass-fail have not been encouraging. A recent survey of schools using pass-fail showed that administrators and teachers were not satisfied with that system of evaluation. Historically, one-fourth of the schools who have attempted pass-fail have abandoned it.

**Advantages.** Pass-fail has the potential to reduce anxiety, eliminate cheating, and reduce destructive competition. It is also possible that the learning environment may be better because students may feel freer to explore the subject in their own way. Furthermore, the pass-fail method may cause students to be less constrained to agree with the teacher. It can be argued that students studying under the pass-fail system still do plenty of work, because they must meet the instructor's minimum standard for passing the course.

**Disadvantages.** Pass-fail is a system which amounts to blanket grading since all students eventually may receive a grade of "pass." Freed from the pressures that letter grades provide, the typical student will work less than usual.

It is particularly difficult to determine the level of work expected for a student to obtain a passing grade. In many cases, there

may be a large variation in the standards for passing from one course to another.

The evaluation that students receive in letter grading provides them with some useful feedback. Since all students receive the same grade and since they are usually not informed concerning their progress, students receive little helpful feedback in this system.

When students are in danger of failing a course, the pressures associated with pass-fail are at least as great as the pressures associated with regular grading procedures.

## Equal Grading

Some instructors with an extremely humanistic educational philosophy believe strongly that all students should be treated equally. This belief extends to grading and these instructors give all of their students identical grades.

Usually this system of grading begins by the teacher making an announcement to the students that everyone who completes the required work satisfactorily will receive a specified grade. The grade most often specified is "B." However, some instructors may elect to give all of their students "A" and others may give every student a "C."

In extreme cases, the system can be used with no stipulations made concerning the quality of work to be completed. For example, a college instructor may announce to his class that everyone enrolled will receive a "B." Usually such extreme evaluation measures are utilized as a protest against the requirement of schools that instructors must assign grades.

Since all students receive the same grade, competition and anxiety should be reduced. There can be an improved atmosphere for learning.

However, since there is no conscious attempt to evaluate students, they receive none of the helpful feedback that the evaluation process can provide. The grades students receive when an instructor uses this method do not provide any information that can be used to distinguish the good students from the poor students.

Frequently, using this system violates the written grading policies of the school so that educators who use this method take a decided risk. Fellow teachers usually do not approve of assigning blanket grades to students. The students themselves (particularly the good students) consider this method to be a coward's way out and often feel that it is an unfair evaluation system.

It can be argued logically that students will not work as hard when they know they are all going to receive the same grade.

Unless the blanket evaluation procedure is carried to extremes, it is still possible for students to attain failing grades if the instructor considers the quality of their work to be unsatisfactory.

In summary, equal grading involves assigning the same grade to all students regardless of the quality of their work. Although equal grading could possibly create a less anxious learning environment, it is risky to attempt it. Instructors who use it are often unpopular with students and fellow teachers.

## Performance Based Evaluation

For some time now, there has been a growing trend to evaluate students on the basis of their mastery of behavioral objectives. Evaluation reports provide students and parents with a detailed account of the objectives the student has mastered. The evaluation of the student by the teacher states in precise terms what concepts have been mastered.

In performanced based evaluation, there usually are no grades. Students merely demonstrate that they are able to master certain instructional objectives.

Needless to say, this type of evaluation requires the student to undergo a new type of learning. Instead of trying to attain a broad, global view of the subject, the student will now find it to his advantage to master the individual skills that are needed to develop proficiency in the subject. Instead of broad cognitive knowledge, this method of grading emphasizes performing certain tasks to demonstrate mastery of small phases of instruction.

Performance based evaluation is particularly suitable for accountability considerations. Teachers cannot be held accountable for the letter grades their students achieve, but they can be held accountable for the instructional objectives their students master.

Figure 5-5 is an example of a part of a performance based evaluation report.

ENGLISH/SOCIAL STUDIES 7

Name _____ Section _____

Progress Report

Map and Globe Skills

unit synopsis

Pupils are called upon to read and interpret maps of increasing complexity. They are asked to make inferences and draw logical conclusions from the data presented on a map or series of maps used together. They work with different global views and with several kinds of map projections in small groups and individually.

unit objectives

| | | |
|---|---|---|
| 1. Upon completion of the map and globe skills workbooks, the pupil should be able to explain the difference between a photo and a map of a specific area, giving advantages and disadvantages of each. | ☐ | meeting requirements |
| | ☐ | not meeting requirements |
| 2. Upon completion of the map skills work, the pupil should draw a map of an area to scale including symbols and legend. | ☐ | meeting requirements |
| | ☐ | not meeting requirements |
| 3. Upon completion of the map skills work, the pupil should be able to locate a site on a map given the latitude and longitude. | ☐ | meeting requirements |
| | ☐ | not meeting requirements |
| 4. Upon completion of the map and globe skills work, the pupil should be able to explain the difference between physical and cultural features on a map. | ☐ | meeting requirements |
| | ☐ | not meeting requirements |
| 5. Upon completion of the map and globe skills work, the pupil should be able to construct agricultural, rainfall, relief, and political maps of a given area. | ☐ | meeting requirements |
| | ☐ | not meeting requirements |
| 6. Upon completion of the map and globe skills work, the pupil should be able to construct a contour map of a given area. | ☐ | meeting requirements |
| | ☐ | not meeting requirements |

Comments:

Figure 5-5
A Performance Based Evaluation Report

## Chapter 6

## ALTERNATIVES TO GRADE POINT AVERAGE AND CLASS RANK

For many years, the calculation of grade point average (G.P.A.) and rank in graduating class has been a standard procedure in almost every secondary school in the United States. The computation of each student's academic position in relation to that of the other members of his class is usually performed to determine the absolute measure of academic proficiency known as grade point average. From these computations, students are ranked from high to low on the basis of their grade point average and this ordinal ranking is reported as class rank.

Thus, a student with an average grade of B in high school has a 3.0 grade point average on a scale that equates an A to 4.0 points, a B to 3.0 points, etc. The same student's grade point average, when compared to the G.P.A. of other members of his class, might allow school officials to conclude that the student ranks 86th in a graduating class of 420. The lowest numbered ranks indicate the highest G.P.A.'s. The student who graduates with the highest G.P.A. is ranked first in the graduating class and is often designated as class valedictorian. The student who graduates with the second highest G.P.A. is ranked second and is often designated as salutatorian.

### Reasons for Computing G.P.A. and Class Rank

**College admission.** Since most applications for admission to college require that the student report his class rank, most high school administrators have believed that it was necessary to calculate class rank so that a college could be informed of the class standing of an applicant for admission.

During the 1960's, there was a shortage of openings for students in most of the colleges in the United States. The rank in graduating class statistic was frequently used as a criterion for admission or rejection to a college or to a specific program within a college. As an example, a student who ranked 156th in a graduating class of 300 might not be admitted to his state university or to a prestigious private school. However, this class rank might be considered acceptable for enrollment in a state supported college, provided the student did not wish to study a curriculum with higher admission criteria such as a pre-law or pre-medicine program.

A study of U.S. high schools in 1972 conducted by the National Association of Secondary School Principals found that 97% of the high schools reported class rank and that 84% of them would continue to do so even if they felt that colleges did not need the statistic for the admissions decision-making process.

However, another survey of colleges determined that only 6 of 617 colleges would deny students admission if their rank in class were not reported and that 90% of the colleges in the United States reported that the absence of grades, G.P.A.'s and class rank would not be a handicap to a candidate's chances for admission.

**Predicting success in college.** In addition to using class rank and G.P.A. for admission decisions, many colleges use either G.P.A. or class rank or both to predict the degree of success that their entering students may expect to achieve. It is a fairly easy procedure to calculate the correlation coefficient between high school G.P.A. and class rank and college G.P.A. or class rank. This correlation can be used along with other data and information to determine the probability of success of an entering student in any academic program. A lot of effort and talent has been utilized in the development of the science of prediction. The prediction of academic success in college has become a rather precise, although far from perfect, science.

Class rank and G.P.A. are not the only predictors of academic success. Other predictors are the results of standardized tests, such as the College Board Examination (CEEB) and the Scholastic Aptitude Test (SAT), the subjective analysis of interviews with college admission officials, and interpretation of letters of recommendation. Frequently several of these criteria are combined in order to obtain an equation which can be used for predicting success in college performance.

The worth of class rank and G.P.A. for prediction of academic success has been studied for many years and it has been found to be a meaningful statistic for that purpose. The results of many correlational studies have shown agreement that *either* class rank *or* G.P.A. is the best single predictor of college success that is known today.

**Incentive to students.** A high grade point average is considered by many educators to be a motivating factor. Although there may be some doubts about the quality of motivation that the desire to improve class rank and G.P.A. provides, it is undoubtedly true that many students, particularly the more able ones, are conscious of their academic standing and are willing to work hard to keep a good G.P.A. or to improve their class rank.

It is also true that the computation of G.P.A. and the determination of class rank give scholars a kind of recognition that has often been denied to them. The recognition of scholarship in many schools has been and is overshadowed by recognition given for athletics and extra-curricular activities.

Many school officials believe that the value of class rank is its motivating capability rather than the information it provides to colleges.

**Determination of recipients of scholarships and financial aid.** Class rank has often been used to determine whether a student could be

awarded a scholarship. NCAA Rule 1.6 prescribes class rank as the criterion to determine the elegibility of athletes for scholarships.

## Criticisms of Ranking and G.P.A.

**Humanistic considerations.** The criticisms of class rank and G.P.A. by existentialists and so-called educational humanists are similar to those they level at grading procedures generally. The critics of ranking find it to be a discriminatory practice. They find that determining class rank and G.P.A. constitutes a "labeling" process.

There is some truth in both of these allegations. The rank in class statistic can be one that will benefit or handicap the student for many years, particularly in the early formative stages of his career. Certainly anyone who has carefully attended to the conversations of college students is aware that they often know the G.P.A.'s of their peers almost as well as they know their names.

**Reliability considerations.** Those who favor abolishing ranking procedures find the G.P.A. to be a useless and irrelevant statistic. Any student's class rank and G.P.A. can vary greatly from school to school, from teacher to teacher, and can be affected by the depth and intensity of the courses the student elects to complete. Since the G.P.A. and class rank are based on grades, which are the subjective judgments of teachers, class rank is a rather unscientific statistic. Furthermore, the grading procedure is steeped in human error and critics of class rank feel that grades constitute an evaluation of the teacher's attitudes toward students rather than a measurement of academic competence.

## Alternatives to G.P.A. and Class Rank

**Variations in ranking and G.P.A.** Frequently, a high school may attempt to compensate for differences in rigor of various curricula by weighting the grade point value of certain courses. Often honors courses and extra courses are considered to be of extra grade point value. Also, a grade in a repeated course may erase or otherwise compensate for an earlier low grade a student received the first time he completed a course.

As an example, suppose a student is taking an honors course in social studies, completing an extra advanced course in biology at a local college after school hours, and repeating a course in typing in an attempt to improve the failing grade received on the first attempt in the class.

Assume the student received all B's. The social studies honors course and the advanced placement biology course grades would each count 3.5 points toward the student's average rather than the 3.0 points of a B in a regular class. The B in typing would count 3.0 points toward the total but would replace the 0.0 points the student received when he failed the course.

**Estimated rank.** Some schools attempt to inform colleges and prospective employers of approximately how a student ranks in relation to other students in his class. This ranking estimate is entirely subjective and usually reflects an opinion of the student's counselor, rather than an actual computation of the student's G.P.A. or class standing.

This estimate may be an educated guess on the part of the counseling staff. It is also possible that the high school may have actually computed the student's exact class rank but has elected to report class standings in percentile ranges. In either case, the student's class standing would be reported as being in the second-highest one-tenth. Rather than reporting that a student has been either estimated or calculated to be 250th in a class of 300, the school would simply report the student's class standing as being between the percentile ranks of 10 and 20. This indicates that the student ranks in the second-lowest one-tenth of his graduating class.

Although estimated class rank does provide some indication of the quality of work in relation to others in the graduating class, it is not as precise or as objective as actual class rank.

Other schools may have administrators that consider grades to be important as motivation agents, but find the computation of G.P.A.'s to be tedious and unscientific. In this case, the high school usually submits the student's full transcript and it then becomes the responsibility of the college admissions office to analyze it for purposes of admission decisions.

The colleges themselves have a minor problem in evaluating the work of students who have attended schools where no class ranking has been made. In colleges where admission is competitive, the alternative to class rank is often to consider scores on standardized tests such as the CEEB, ACT, or the SAT.

The standardized tests are more objective, more reliable, and less subject to human error than class rank and G.P.A. However, to substitute tests for class rank as the criterion for college admission amounts to substituting a one-morning test performance for the student's four year academic record.

Furthermore the test-wise student is given an unfair advantage when decisions are based solely on the results of standardized tests. However, national norms are available for standardized tests, which enables school officials to make a meaningful comparison of students who attend different schools. Such a comparison is, of course, not possible with class rank or G.P.A.

## Summary

Class rank and grade point averages are computed by most U.S. high schools. School officials consider them to be statistics which are

valuable in predicting success in college. The desire to increase one's G.P.A. or class rank is probably a strong motivating force for many students.

Critics of class rank and G.P.A. consider these statistics to be erroneous and meaningless. They feel that class rank and G.P.A. label students.

Although the computation of class rank is no longer a necessity for admission decisions in most colleges, it is a valuable statistic in admission decisions for selective colleges and for competitive programs.

When a school eliminates class rank and the G.P.A., there are several possible results. First, some of the pressure for grades motivators will be eliminated. Second, some of the students will no longer try to take easy courses in order to improve their G.P.A. Third, perhaps some of the pressure on children to receive "good" grades will be removed and some of the pressure will be removed from teachers and administrators. Fourth, the guidance personnel and school administrators will need to work out an acceptable method for informing college admissions offices of the potential ability of students who attended a school that did not find it advantageous to compute G.P.A. or class rank.

Additional copies of this bulletin may be ordered from:

CURRICULUM RESEARCH AND DEVELOPMENT CENTER

Jamison Hall
School of Education
Indiana State University
Terre Haute, Indiana 47809

Also available:

INDIANA MIDDLE SCHOOL SURVEY

A DEVELOPING CURRICULUM IN ENGLISH
Kindergarten through Grade Twelve

ROLE AND STATUS OF CURRICULUM WORKERS IN INDIANA

ROLE AND STATUS OF CONTENT, MEDIA, AND SPECIAL SERVICES SUPERVISORS

CONSERVATION EDUCATION IN INDIANA SCHOOLS

ECOLOGY, CONSERVATION, POLLUTION
A bibliography of materials

LEARNING ACTIVITY PACKAGES

OPEN SPACE SCHOOLS
A bibliography of materials

USING THE BLOCK PROGRAM TO PREPARE ELEMENTARY SCHOOL TEACHERS AT ISU

ACTIVITIES FOR ELEMENTARY SCHOOL MATHEMATICS ENRICHMENT

LITERATURE AND LEARNING TO READ

ORAL LANGUAGE AND READING